A Study of Quasar Clustering at $z > 2.7$ from the Palomar Transit Grism Survey

Andrew W. Stephens [1] and Donald P. Schneider [2]

Department of Astronomy & Astrophysics, Pennsylvania State University, University Park, Pennsylvania 16802

Maarten Schmidt

Palomar Observatory, California Institute of Technology, Pasadena, California 91125

James E. Gunn [3]

Princeton University Observatory, Princeton, New Jersey 08544

David H. Weinberg

Department of Astronomy, Ohio State University, Columbus, Ohio 43210

ABSTRACT

The quest for structure indicators at earlier and earlier times in the evolution of the universe has led to the search for objects with ever higher redshifts. The Palomar Transit Grism Survey has produced a large sample of high redshift quasars ($z > 2.7$), allowing statistical analysis of correlation between quasar positions. In this study, clustering is identified through comparison with 100 000 Monte Carlo generated, randomly populated volumes, which are identical to the observed region in spatial coordinates, redshift distribution, and number of quasars. Three pairs have been observed with comoving separations of 11.34, 12.97, and 24.13 h_{50}^{-1}Mpc (assuming $q_0 = 0.5$), smaller separations than would be expected to arise by chance in an unclustered distribution. Selection effects are ruled out as a false source of clustering by scrambling the observed quasar coordinates and redshifts, which gives a pair separation distribution nearly identical to that of the Monte Carlo distribution. Tests using the distribution of pair separations and nearest neighbor distances show that the observed pairs have a probability less than 0.1% of arising in an unclustered distribution. Using a maximum likelihood technique to estimate the correlation length r_0, assuming $\xi(r) = (r/r_0)^{-1.8}$, we find $r_0 = 35 \pm 15\ h_{50}^{-1}$Mpc (comoving, $q_0 = 0.5$, 1σ errors), a value much larger than the correlation length of present-day galaxies.

1. Introduction

The cosmic microwave background is remarkably uniform, but redshift surveys of the local galaxy distribution reveal coherent structures that extend over 100 h_{50}^{-1}Mpc or more ($h_{50} \equiv H_0/50$ km s^{-1} Mpc^{-1}). One would like to trace the growth of this structure from the very early epoch probed by the microwave background through to the present day. This objective has motivated clustering studies of homogeneous quasar samples, an approach pioneered by Osmer (1981). In this paper, we analyze the clustering of 56

(of 90) high-redshift quasars ($2.7 < z < 4.75$) from the Palomar Transit Grism Survey (PTGS)(Schneider, Schmidt, and Gunn 1994, hereafter SSG), characterizing structure in a sample that contains some of the most distant quasars discovered to date.

When Osmer (1981) first obtained a survey of 174 quasars ($z < 3.5$), he was faced with the problem of how to quantify clustering, which he defined as density enhancements significantly in excess of expected random fluctuations. After considering many statistical methods, he adapted and applied binning analysis, the nearest neighbor test, and the correlation function to his sample. These tests dismissed several close groups picked out by eye as insignificant, and Osmer claimed no clustering on the 100–3000 h_{50}^{-1}Mpc scale, which was later confirmed by Webster (1982) using Fourier Power Spectrum Analysis.

Concerned that the lack of clustering evidence was an observational result of previous small sample sizes, Shaver (1984) devised a method of detecting clustering by comparing quasars close both in redshift and position to those with large differences in either redshift or position. This method allows for the analysis of inhomogeneous quasar samples, since selection effects cancel out. He applied his method to the ~ 2000 quasars in the Véron catalog (Véron-Cetty & Véron 1984), revealing possible clustering similar to that of galaxies of the current epoch, which was later confirmed and quantified by Kruszewski (1988), who detected statistically significant clustering on scales up to 32 h_{50}^{-1}Mpc at low redshift but none at $z > 1.3$.

Later studies of quasar clustering attempted to achieve large sample sizes, while still maintaining homogeneity to minimize selection effects. Shanks et al. (1987) used 170 quasars from UK Schmidt plates, complete to $z = 2.2$ and $B < 21$, and detected clustering on small scales ($r < 20$ h_{50}^{-1}Mpc) but not on large scales ($20 < r < 2000$ h_{50}^{-1}Mpc). Drinkwater (1988) found no clustering on scales of 20–200 h_{50}^{-1}Mpc in a sample of 862 quasars ($1.8 < z < 2.6$) from six UKST low dispersion objective prism plates. Crampton, Cowley & Hartwick (1989) discovered an unusually large group of 23 quasars at $z \sim 1.1$, but no statistical evidence for clustering in the remaining 91% of the Canada-France-Hawaii Telescope grens survey. Such large quasar groups (LQGs) have been studied by Komberg et al. (1996), who detected 12 LQGs with redshifts $0.5 < z \leq 2$, and sizes from ~ 140 to ~ 320 h_{50}^{-1}Mpc.

Recently Shanks and Boyle (1994) have examined the clustering in three combined surveys with a total of 690 QSOs at $0.3 < z < 2.2$, claiming a $\sim 4\sigma$ detection with a best value of $\bar{\xi}$(r< 20 h_{50}^{-1}Mpc)=0.93 ± 0.34 for the mean 2-point correlation function of pairs with comoving separation r< 20 h_{50}^{-1}Mpc (q_0=0.5). Kundic (1997) has also evaluated the clustering properties of 232 quasars from the PTGS and found significant clustering on scales less than 40 h_{50}^{-1}Mpc.

This study will concentrate on the high-redshift quasars ($2.7 < z < 4.75$) from the PTGS. The extreme distance of these quasars allows analysis of the universe when it was only about ten percent of its current age, providing the earliest observational link between the homogeneity of the universe at recombination, and the clumpiness of today's environs. If quasar clustering is indicative of the large-scale distribution of their parent galaxy groups, quantifying clustering of these very "young" objects could help discriminate between various models of structure formation.

2. Observations

The PTGS was carried out with the Hale telescope, using the 4-Shooter camera/spectrograph (Gunn et al. 1987) in time delay-and-integrate (TDI) mode in order to obtain large numbers of low-resolution slitless spectra. The total effective area of the survey is 61.47 square degrees, which is covered in six narrow

strips 8.5′ wide. A computer algorithm was used to search the ∼600,000 grism spectra for the presence of emission lines. The primary target was the Lyman α line, which appears in the survey bandpass at redshifts between 2.7 and 4.8. A total of 928 emission-line objects were found. Approximately 70% of the objects were low-redshift emission-line galaxies; 90 of the sources were quasars at redshifts between 2.757 and 4.733.

This study of high redshift quasar clustering concentrates on two strips of the survey. The "MN" strip spans right ascensions from 09^h12^m to 16^h45^m (with a 4^m break caused by clouds), at a central declination of $+46°27′$, covering 9.98 square degrees and containing 24 quasars with $z > 2.7$. The "QR" strip spans right ascensions from 08^h09^m to 17^h14^m, at a central declination of $+47°39′$, covering 11.91 square degrees, with 32 $z > 2.7$ quasars. These two strips were selected for analysis because they have the faintest effective flux limits for quasar detection (this limit is determined by a combination of seeing, exposure time, sky brightness, and throughput of instrument; see SSG). While they cover but a third of the survey area, they contain nearly two thirds of the high redshift survey quasars (56 out of 90). The remaining strips are too sparse to yield statistically significant results (positive or negative) in our tests, though we include them in our calculation of the correlation length in section 4.6 below.

3. Numerical Methods

To determine absolute distances between quasar pairs, the quasar's equatorial coordinates are converted to rectangular comoving coordinates using the quasar's redshift as a measure of its distance

$$r = \frac{c}{H_0} \times 2 \left(1 - \frac{1}{\sqrt{1+z}} \right), \tag{1}$$

assuming that $q_0 = \frac{1}{2}, (\Omega = 1)$.

To investigate quasar clustering in this sample, the observed sample is compared with a large number of randomly populated Monte Carlo generated samples which match the actual sample in sky coordinates and redshift distribution. The right ascension coordinate is chosen randomly between the upper and lower strip limits, except in the MN simulation. While obtaining the direct imaging data for the MN strip, the passage of a moderately thick cloud through the field rendered the data between right ascensions 11^h34^m and 11^h38^m unusable (see SSG). If the simulation created a quasar with a position in this region, the quasar was deleted and a new one selected to take its place. The declination coordinate is selected through a fractional areas technique which equates a random number to the fraction of the area below a declination, yielding a random declination given by

$$\delta = \sin^{-1}[R(\sin \delta_2 - \sin \delta_1) + \sin \delta_1], \tag{2}$$

where R is a randomly chosen number between zero and one, and δ_1 and δ_2 are the declination limits of the test strips. Redshifts are chosen randomly from a smooth model of the observed redshift distribution (Figure 1).

All random numbers are generated by the subroutine RAN2 obtained from *Numerical Recipes* (Press et al. 1992), which was tested and found to be sufficiently random over the required number of calls. These methods are used for each quasar in each trial, where the number of quasars randomly generated in each strip matches the observed number: 24 in the MN strip, and 32 in the QR strip. Each strip is numerically
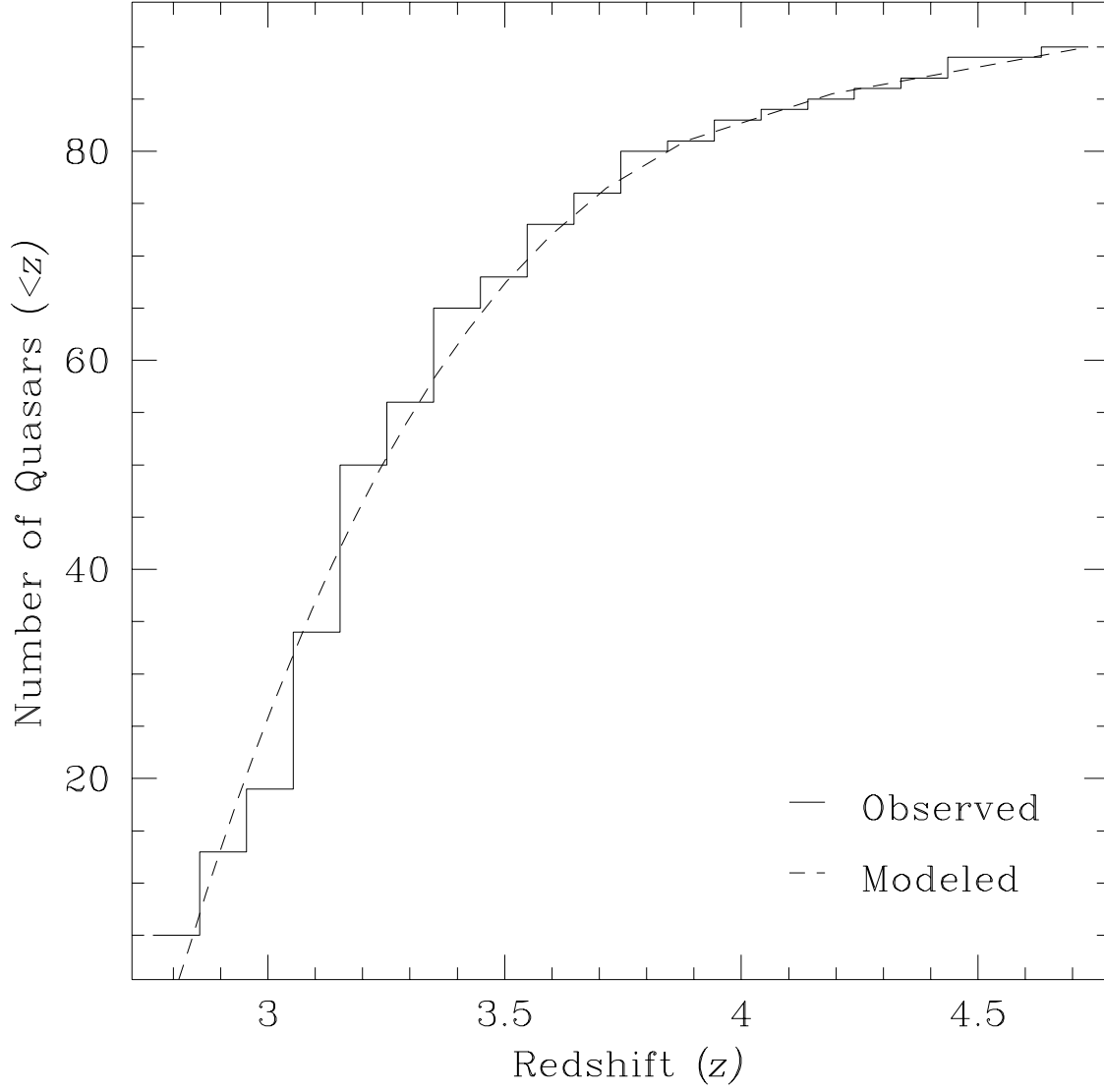
Fig. 1.— Observed (solid) and modeled (dashed) PTGS redshift distribution.

– 5 –

modeled 100 000 times for the following statistical analyses.

4. Tests for Clustering

4.1. Pair Separations

The first comparison is between the observed and average random distributions of separations between quasar pairs. The distance is found between all possible quasars pairs in the volume, and these distances are placed into 25 logarithmic bins, where the bins are created to span nearly the entire range of observed separations, from 10 h_{50}^{-1}Mpc to approximately 8000 h_{50}^{-1}Mpc.

The average random pair separation distribution is found by adding each pair separation to the appropriate distance bin as each trial is executed, yielding the cumulative number of pairs in each bin from the entire 100 000 trials, which is then normalized by dividing by the number of trials. The observed and average random quasar pair separation distributions are plotted for each of the two strips in Figures 2 and 3.

These plots show that for separations greater than ~ 60 h_{50}^{-1}Mpc, the observed distribution of pair separations is very nearly random, while at separations < 30 h_{50}^{-1}Mpc, quasar pairs exist where none should if the distribution were random. The apparent deficit of observed pairs in the 20-100 h_{50}^{-1}Mpc bins of the QR strip is not statistically significant, as only 2.24 pairs are expected. By contrast, the single pair in the smallest bin for this strip is quite significant; only 2.7% of Monte Carlo trials have one or more pairs in this bin. The two pairs with $r < 30$ h_{50}^{-1}Mpc in the MN strip are much more significant still; only 0.523% of the 100 000 Monte Carlo trials have two or more pairs in the first four bins.

4.2. Scrambling

In order to determine if these excess quasar pairs are due to selection effects, the coordinates and redshifts of each of the observed quasars in the survey are scrambled and randomly redistributed 100 000 times, so that all coordinates are assigned to all redshifts in random combinations. The resulting average scrambled pair separation distribution for each strip is also shown in Figures 2 & 3.

For the MN strip, the scrambled pair separation distribution very nearly matches the observed distribution on scales larger than ~ 60 h_{50}^{-1}Mpc, but scrambling does not produce significant numbers of close quasar pairs like the ones observed. Similarly in the QR strip, the scrambled distribution is close to the observed distribution, except for in the smallest bin, where again, no significant number of close quasar pairs are evidenced like the one observed. These results are nearly identical to the Monte Carlo generated results at the clustering scale under investigation, and show that the observed quasar pairs are *not* due to selection effects.

4.3. Minimum Separations

For another test, one can ask how the minimum quasar pair separation in each of the randomly generated strips compares with the observed minimum quasar pair separation. This is answered by Figure 4, which shows the cumulative probability distribution for any minimum separation in each quasar strip.
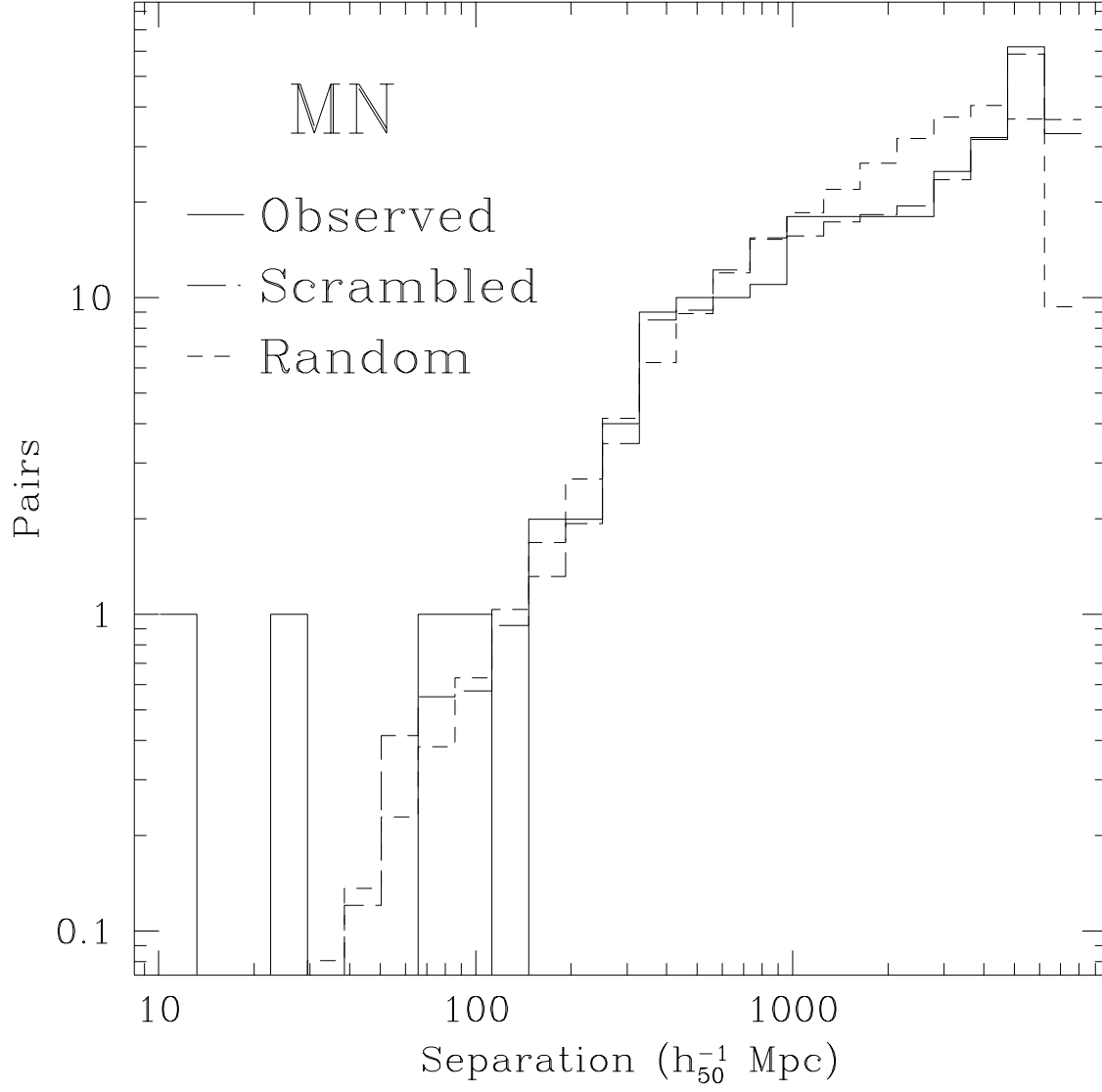
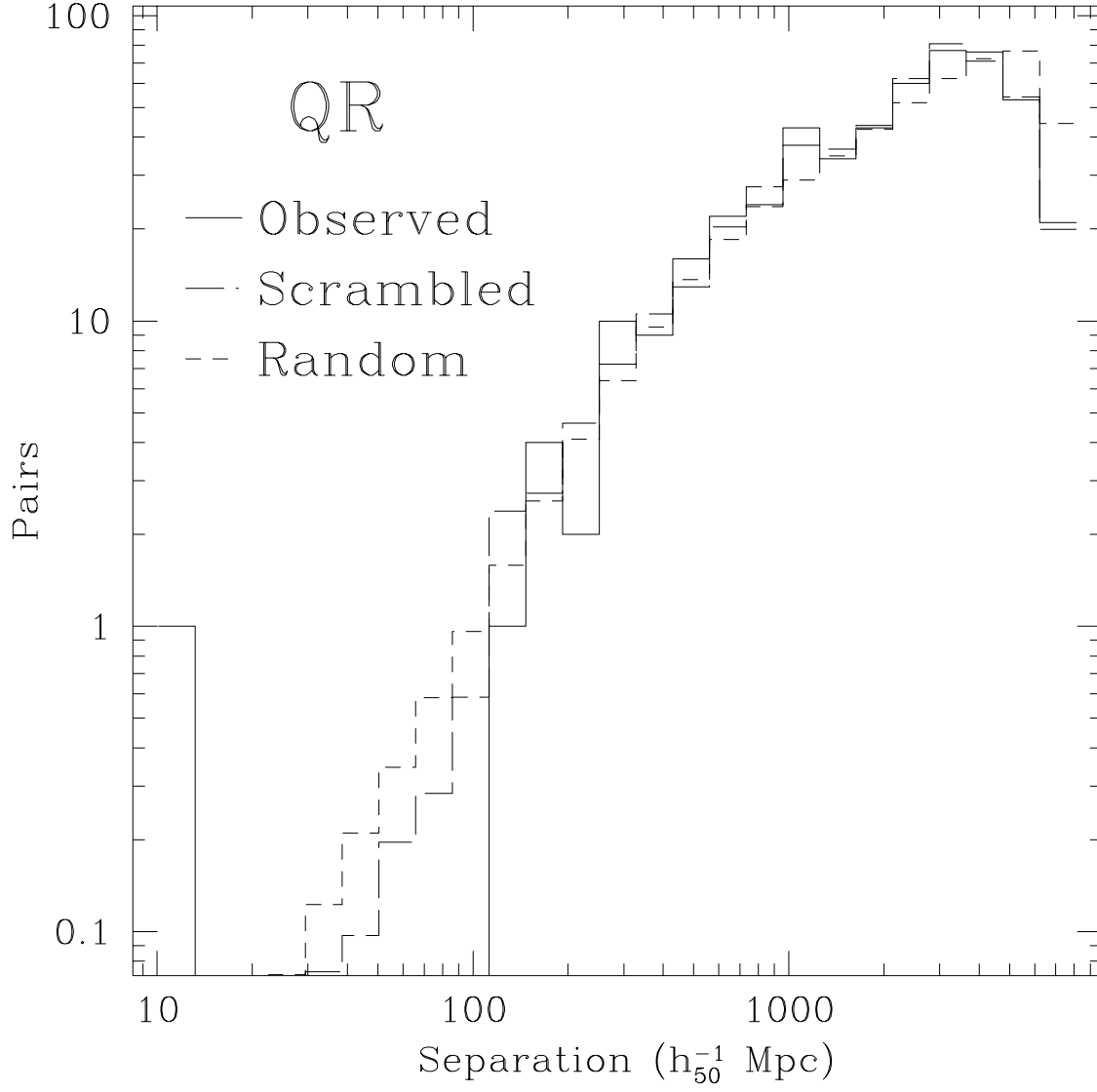Fig. 2.— MN Strip: Observed (solid), Scrambled (long dashed) & Random (short dashed) distributions.

Fig. 3.— QR Strip: Observed (solid), Scrambled (long dashed) & Random (short dashed) distributions.

This distribution is obtained by determining the distance between the closest quasar pair in each of the 100 000 trials, placing them cumulatively into the standard logarithmic bins, and normalizing to one by dividing by 100 000. Superimposed onto their respective curves are filled circles representing the closest observed quasar pairs for each strip, and the third quasar pair (with a separation of 24.13 h_{50}^{-1}Mpc) for reference. This plot shows that the probability of obtaining *one* minimum separation as small as that exhibited in any of these three pairs is less than 10%.

4.4. Nearest Neighbor Test

A nearest neighbor distance (the distance to the closest quasar) is calculated for each quasar in the survey, and these distances are placed into the same logarithmic bins. For the randomly generated quasars, the nearest neighbor distance is calculated for each quasar in each trial, placed into bins, and the bins are averaged over the 100 000 trials, yielding the mean number of quasar pairs in each bin. This is compared to the observed number of nearest neighbor pairs in each bin in Figures 5 & 6 for the MN and QR strips. From these plots it is evident that for the MN strip, the nearest neighbor separation distribution peaks at $\sim 300\ h_{50}^{-1}$Mpc, both for the randomly generated quasars and the observed quasars, *but* the observed nearest neighbor distribution is dramatically higher at very small separations ($< 30\ h_{50}^{-1}$Mpc), where the two quasar pairs lie (the plot registers two in each bin because a nearest neighbor distance was measured for *each* quasar, hence a single close pair will produce two identical nearest neighbor distances). Both the observed and the random distributions for the QR strip also peaks near $\sim 300\ h_{50}^{-1}$Mpc, but here again the observed distribution is elevated at the smallest distance bin, corresponding to the single pair in the QR strip.

4.5. Probabilities of Observed Pair Separations

Assuming a Poisson distribution for the occurrence of a close quasar pair, it is possible to calculate the probability of such a pair. The probability of obtaining x pairs in any bin in the absence of clustering is

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}, \tag{3}$$

Table 1.   Probabilities of obtaining pairs – MN Strip

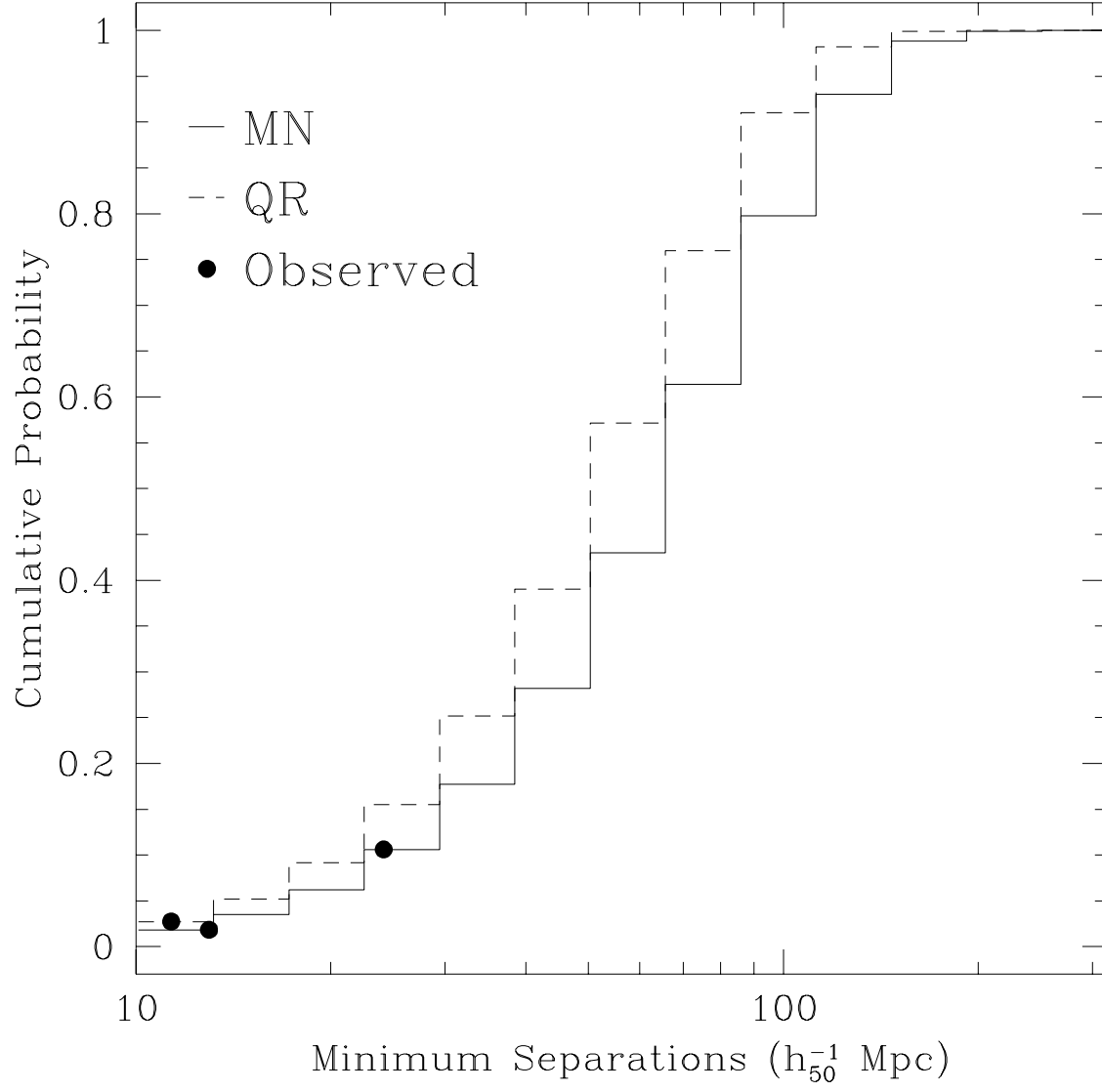| Pairs	Pair Separations in h_{50}^{-1}Mpc							
	11.54	15.08	19.71	25.76	33.66	44.00	57.50	75.16
0	98.158	**98.310**	**97.226**	95.319	**92.030**	**87.320**	**79.563**	68.193
1	**1.825**	1.675	2.735	**4.570**	7.644	11.840	18.190	**26.106**
2	0.017	0.014	0.038	0.110	0.317	0.803	2.079	4.997
3	0.000	0.000	0.000	0.002	0.009	0.036	0.158	0.638
4	0.000	0.000	0.000	0.000	0.000	0.001	0.009	0.061

Fig. 4.— Minimum random quasar pair separations; filled circles indicate observed quasar pairs.
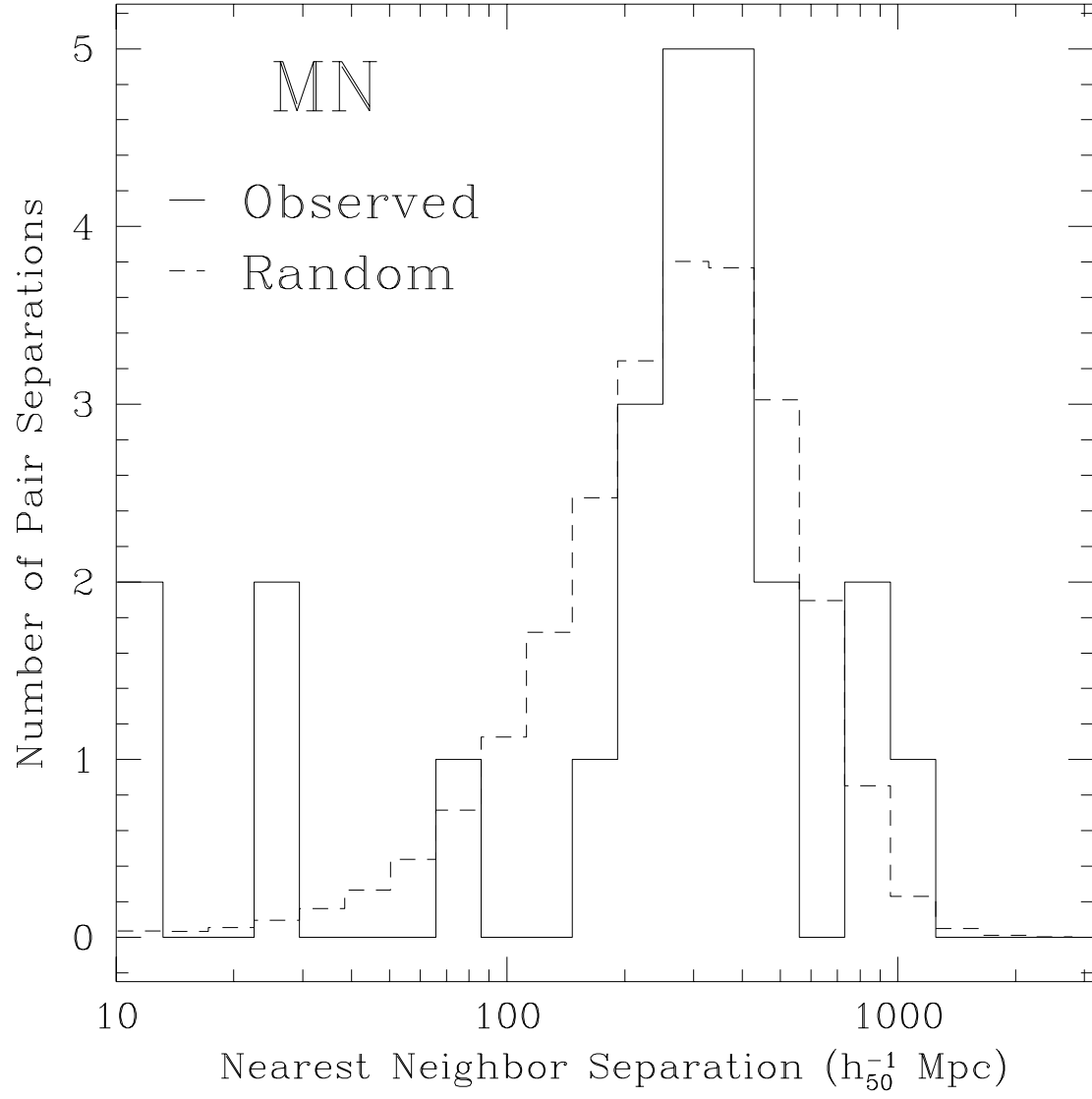
Fig. 5.— Observed (solid) and mean random (dashed) nearest neighbor distribution of the MN strip.
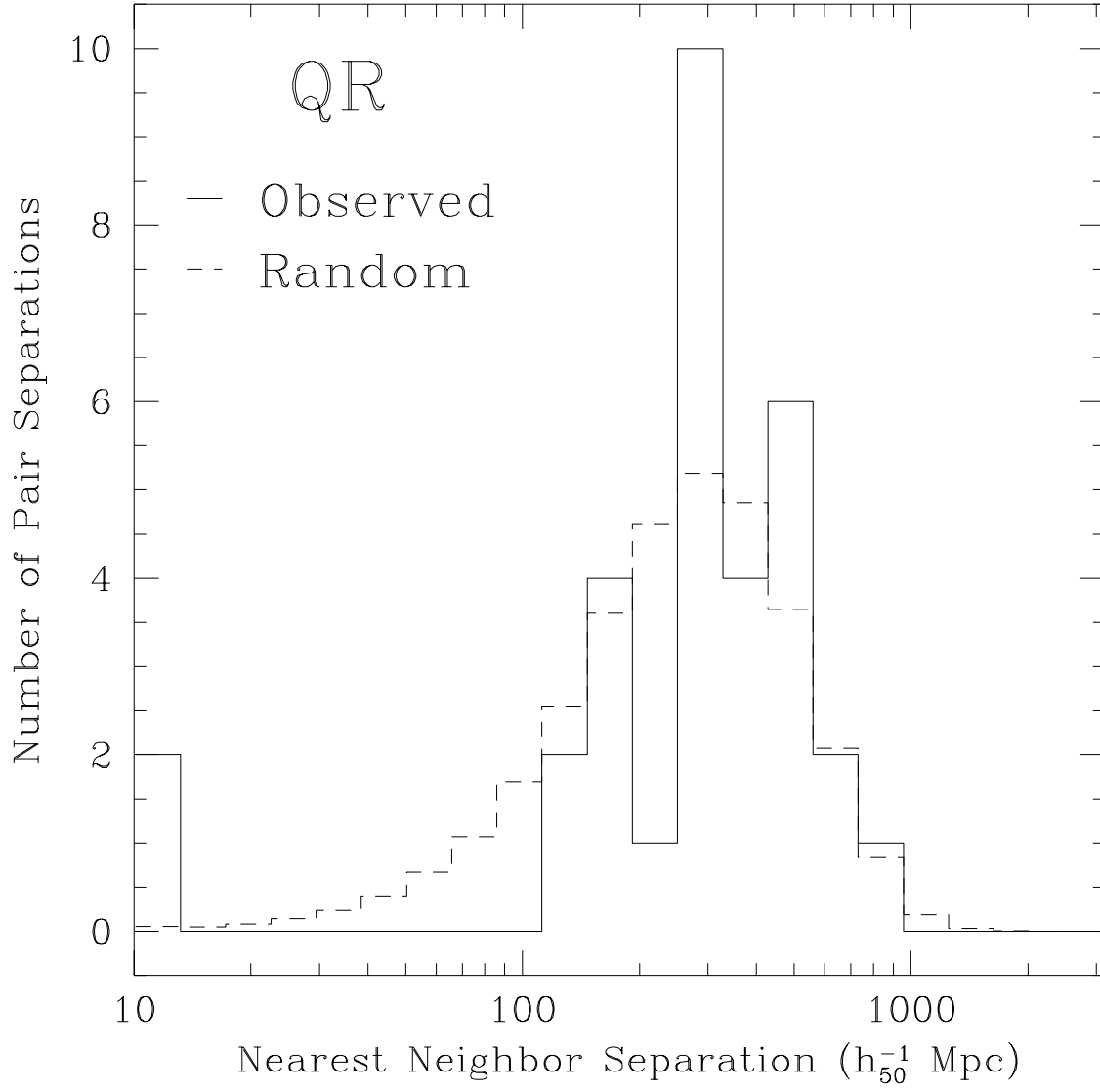
Fig. 6.— Observed (solid) and mean random (dashed) nearest neighbor distribution of the QR strip.

where μ is the mean number of pairs in that bin taken from the Monte Carlo simulations. These probabilities are given as percentages for the first eight bins in Tables 1 and 2. The probability corresponding to the observed number of quasar pairs in each bin is highlighted, revealing that the first and fourth bins of the MN strip, and the first bin in the QR strip are virtually the only bins which do *not* conform to the predicted most probable number of pairs per bin.

4.6.   Correlation Function

The most widely used descriptor of clustering is the correlation function

$$\xi(r) = \frac{N_{obs}(r)}{N_{ran}(r)} - 1, \tag{4}$$

where $N_{obs}(r)$ is the observed number of pairs with separation r, and $N_{ran}(r)$ is the number of pairs expected for a random, unclustered distribution. Thus, an unclustered distribution should yield $\xi(r) = 0$ on average. With only three small-separation pairs, our estimate will necessarily be quite uncertain, but it is useful to present our results in this form to allow comparison with other estimates.

The histograms in Figures 7 & 8 show $1 + \xi(r)$ estimated in logarithmic bins using equation (4). The long-dashed line shows the expectation $\xi(r) = 0$ for an unclustered distribution. The dotted lines indicate the one-sigma envelope for an unclustered distribution due to Poisson fluctuations in the number of pairs, where the standard deviation for each bin is determined from the distribution of quasar pairs from the Monte Carlo simulations,

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2}, \tag{5}$$

where $\bar{x}$ is the mean number of pairs in the bin, and x_i is the number in the ith Monte Carlo simulation. The pair counts at large separations are consistent with Poisson fluctuations in an unclustered distribution, but the high estimates of $\xi(r)$ at $r < 30\ h_{50}^{-1}\mathrm{Mpc}$ are clearly inconsistent with a random quasar distribution.

Table 2.   Probabilities of obtaining pairs – QR Strip

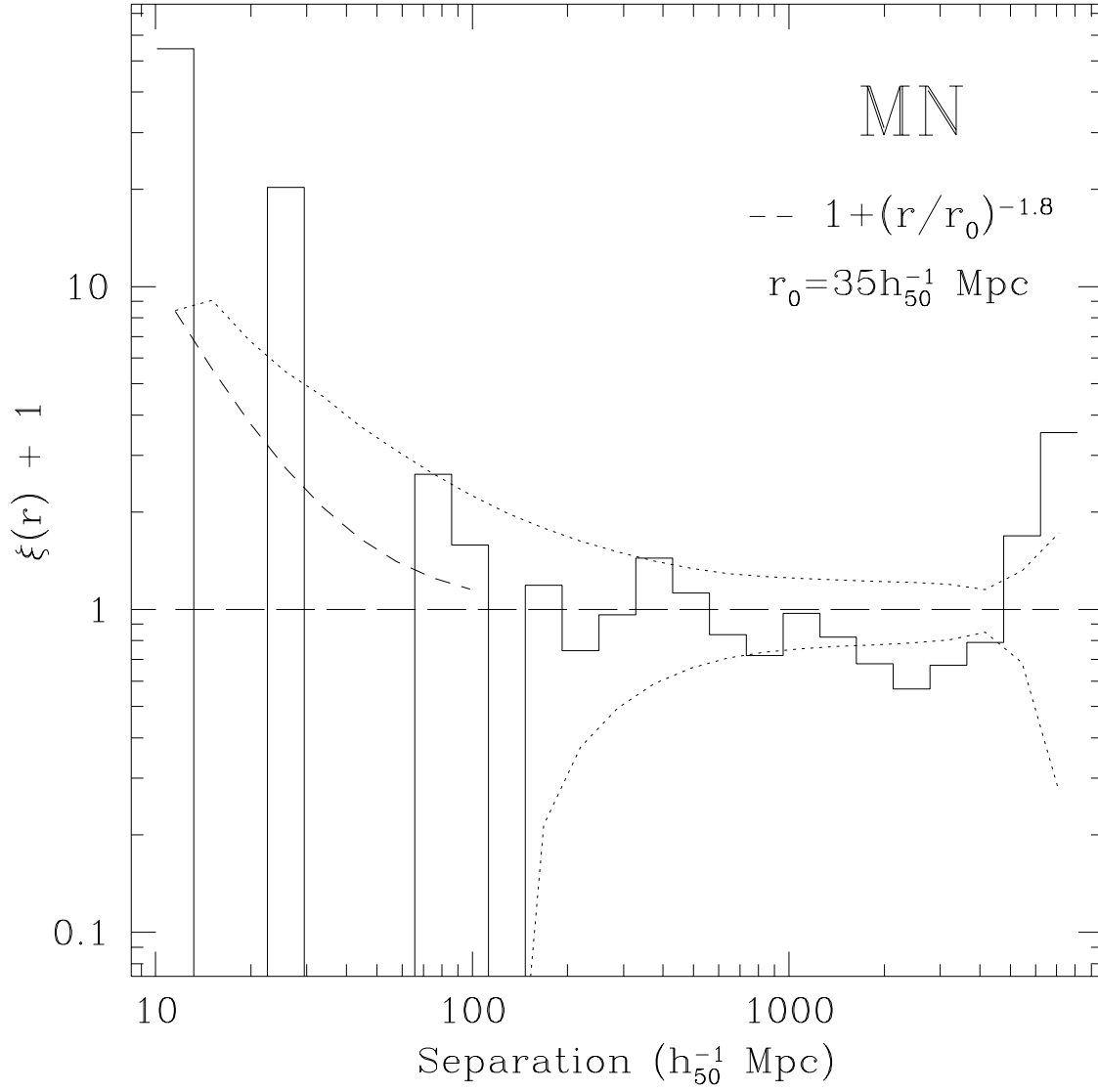| Pairs	Pair Separations in $h_{50}^{-1}\mathrm{Mpc}$							
	11.54	15.08	19.71	25.76	33.66	44.00	57.50	75.16
0	97.269	**97.452**	95.846	**92.986**	**88.649**	**81.413**	**70.484**	**56.109**
1	**2.693**	2.515	4.067	6.762	10.681	16.742	24.654	32.424
2	0.037	0.032	0.086	0.246	0.643	1.721	4.312	9.368
3	0.000	0.000	0.001	0.006	0.026	0.118	0.503	1.805
4	0.000	0.000	0.000	0.000	0.001	0.006	0.044	0.261

Fig. 7.— Correlation function + 1 for the MN strip, the corresponding one sigma error envelope (dotted), and the predicted $\xi(r) + 1$ assuming $r_0 = 35\ h_{50}^{-1}$Mpc.
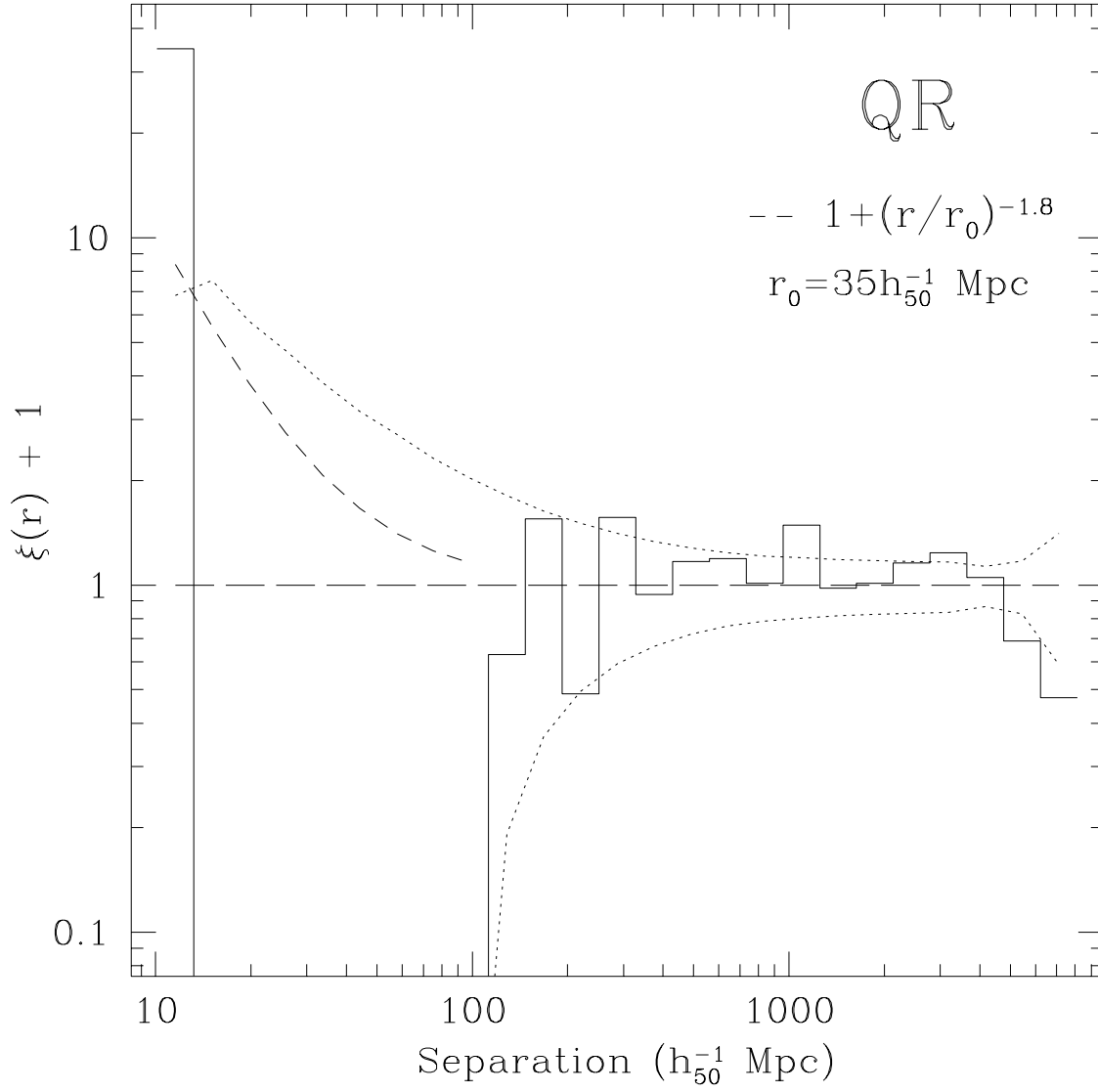
Fig. 8.— Correlation function + 1 for the QR strip, the corresponding one sigma error envelope (dotted), and the predicted $\xi(r) + 1$ assuming $r_0 = 35\ h_{50}^{-1}$Mpc.

Because the quasar distribution is sparse, the probability of obtaining N_{obs} pairs when N_{ran} are expected is given by Poisson statistics (equation [3]). This makes it simple to devise a maximum likelihood procedure for estimating parameters of the correlation function (Croft et al. 1997; Popowski et al. 1997). Because our sample is so small we will assume that the true quasar correlation function has the same *form* as the present-day galaxy correlation function

$$\xi(r) = (r/r_0)^{-1.8}, \tag{6}$$

and estimate only the correlation length r_0 itself.

The calculation is performed simultaneously on all six strips of the PTGS. Using 1 h_{50}^{-1} Mpc bins, we sum the observed and random values of each strip for each bin, then find the value of r_0 that maximizes the likelihood of obtaining the observed number of pairs. Figure 9 shows the results of this analysis, giving the maximum likelihood estimate $\hat{r_0}$ using the first x bins (each 1 h_{50}^{-1}Mpc). As the number of bins included in the calculation is increased, the number of quasar pairs increases sporadically. With each new pair, the estimate $\hat{r_0}$ jumps, then declines smoothly until the next observed pair is encountered.

The 1σ error in r_0 can be calculated from the second derivative of the likelihood function at its maximum,

$$\sigma = \left(\frac{\partial^2 \ln L}{\partial r_0^2}\right)^{-1/2}. \tag{7}$$

A more conservative error range, corresponding to $\sim 2\sigma$ for Gaussian statistics, is given by finding the values of r_0 for which the likelihood falls to 10% of its peak value,

$$\frac{L(r_0)}{L(\hat{r_0})} = 0.1. \tag{8}$$

The dashed and dotted lines in Figure 9 show these error ranges. The estimate of r_0 depends on the number of bins we include in the calculation, but it is fairly stable for large bin numbers at $r_0 = 35 \pm 15 \ h_{50}^{-1}$Mpc (1 σ errors).

Our maximum likelihood estimate of r_0 is much larger than the value of $r_0 \sim 13 \ h_{50}^{-1}$Mpc estimated by Shanks & Boyle (1994), though it is just consistent with this value at the 2σ level. Given the small number of quasars in the present sample and the large resulting uncertainty in r_0, a statistical fluke cannot be ruled out. Alternatively, the unusually luminous quasars in this higher redshift sample may reside in environments whose clustering is strongly "biased" with respect to the underlying mass distribution, e.g., at the very highest peaks of the primordial density field.

5. Conclusion

Evidence has been presented to substantiate high redshift quasar clustering on scales less than 30 h_{50}^{-1}Mpc (comoving). Through 100 000 Monte Carlo simulations, the observed quasar separation distribution is found to match a purely random distribution *except* for bins smaller than 30 h_{50}^{-1}Mpc, where three observed pairs are located. By scrambling and randomly redistributing the observed quasar coordinates
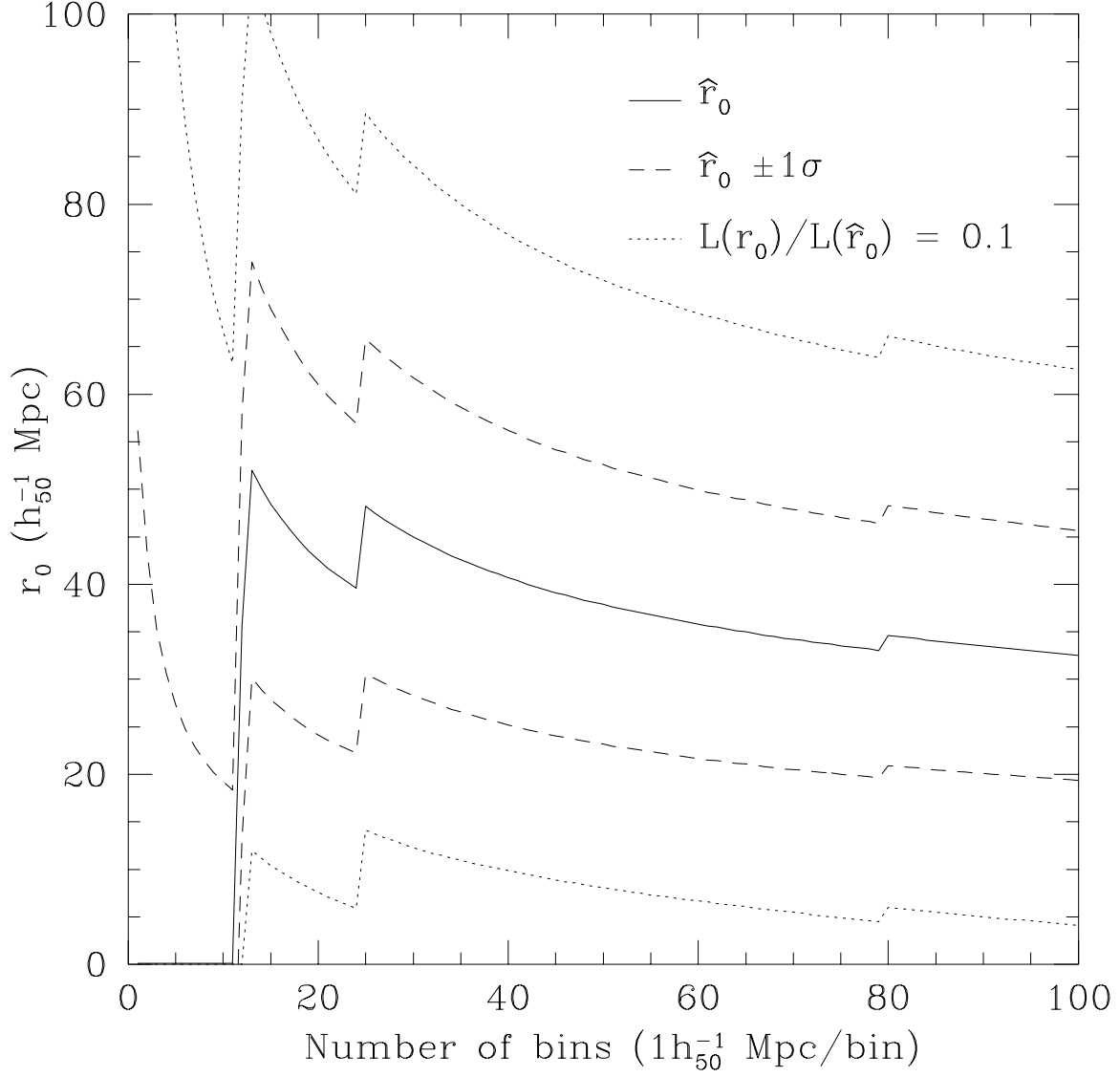
Fig. 9.— Maximum likelihood estimate of the correlation length r_0 as a function of the number of 1 h_{50}^{-1}Mpc bins included in the calculation (solid line). Dashed lines show the 1σ range on r_0. Dotted lines show the range within which the likelihood exceeds 10% of its peak value.

and redshifts, it is shown that the close quasar pairs are not a result of selection effects, and that the scrambled separation distribution is nearly indistinguishable from the Monte Carlo distribution. Analysis of the minimum quasar pair separations for each of $100\,000$ strip generations reveals that the observed quasar pair separations are significantly smaller than the expected minimum quasar pair separation of 65-75 h_{50}^{-1}Mpc. The nearest neighbor test also tells a tale of high redshift quasar clustering on scales of less than 30 h_{50}^{-1}Mpc, showing spikes in the nearest neighbor distance distribution in both the MN and QR strips, while still matching the random Monte Carlo distribution on larger scales. By assuming a Poisson distribution, the probability of the observed quasar pairs is found to be between 2% - 5% each, while the Monte Carlo simulations show that the probability of finding both pairs in the MN strip with separations ≤ 30 h_{50}^{-1}Mpc is $\sim 0.5\%$. The correlation function $\xi(r)$ is large and statistically significant for $r < 30$ h_{50}^{-1}Mpc in both strips.

Our results are consistent with those of Kundic (1997), who analyzed the full PTGS quasar sample (we have focused on the high-redshift $z > 2.7$ quasars). Kundic (1997) also finds statistically significant clustering in the quasar sample on scales $r < 40$ h_{50}^{-1}Mpc, and he finds marginal evidence for an increase in clustering strength towards high redshift.

Assuming $\xi(r) = (r/r_0)^{-1.8}$, we estimate $r_0 = 35 \pm 15$ h_{50}^{-1}Mpc for these high-redshift quasars, much higher than the correlation length $r_0 \sim 10$ h_{50}^{-1}Mpc of present-day galaxies, or the correlation length $r_0 \sim 13$ h_{50}^{-1}Mpc estimated by Shanks & Boyle (1994) for lower redshift quasars. While the statistical uncertainties from this sparse sample are considerable, our results suggest that the most luminous objects in the high-redshift universe also exhibit unusually strong spatial clustering.

REFERENCES

Crampton, D., Cowley, A. P., & Hartwick, F. D. A. 1989, ApJ, 345, 59

Croft, R. A. C., Dalton, G. B., Efstathiou, W. S., & Maddox, S. 1997, MNRAS, submitted

Drinkwater, M. 1988, MNRAS, 235, 1111

Gunn, J. E., et al. 1987, Opt. Eng., 26, 779

Komberg, B. V., Kravtsov, A. V., & Lukash, V. N. 1996, MNRAS, 282,713

Kruszewski, A. 1988, Acta Astron., 38, 155

Kundic, T. 1997, ApJ, submitted

Osmer, P. S. 1981, ApJ, 247, 762

Popowski, P., Weinberg, D.H., Ryden, B. S., & Osmer, P. S. 1997, in preparation

Press et. al., Numerical Recipes, Cambridge University Press, 1992

Schneider, D.P., Schmidt, M., & Gunn, J.E. 1994, AJ, 107, 1245

Shanks, T., Boyle, B.J. 1994, MNRAS, 271, 753

Shanks, T., Fong, R., Boyle, B. J., & Peterson, B. A. 1987, MNRAS, 227, 739

Shaver, P. A. 1984, A&A, 136, L9

Véron-Cetty, M. P., & Véron, P. 1984, ESO Sci. Rep. No. 1

Webster, A. 1982, MNRAS, 199, 683